

---


12-1-2007

## The Voice Transcription Technique: Use of Voice Recognition Software to Transcribe Digital Interview Data in Qualitative Research

Jennifer L. Matheson

Colorado Stat University, [jemathes@cahs.colostate.edu](mailto:jemathes@cahs.colostate.edu)

Follow this and additional works at: <https://nsuworks.nova.edu/tqr>

 Part of the [Quantitative, Qualitative, Comparative, and Historical Methodologies Commons](#), and the [Social Statistics Commons](#)

---

### Recommended APA Citation

Matheson, J. L. (2007). The Voice Transcription Technique: Use of Voice Recognition Software to Transcribe Digital Interview Data in Qualitative Research . *The Qualitative Report*, 12(4), 547-560. Retrieved from <https://nsuworks.nova.edu/tqr/vol12/iss4/1>

This Article is brought to you for free and open access by the The Qualitative Report at NSUWorks. It has been accepted for inclusion in The Qualitative Report by an authorized administrator of NSUWorks. For more information, please contact [nsuworks@nova.edu](mailto:nsuworks@nova.edu).

---



**Qualitative Research Graduate Certificate**  
*Indulge in Culture*  
Exclusively Online • 18 Credits  
**LEARN MORE**

NSU  
NOVA SOUTHEASTERN  
UNIVERSITY

NOVA SOUTHEASTERN

## The Voice Transcription Technique: Use of Voice Recognition Software to Transcribe Digital Interview Data in Qualitative Research

### Abstract

Transcribing interview data is a time-consuming task that most qualitative researchers dislike. Transcribing is even more difficult for people with physical limitations because traditional transcribing requires manual dexterity and the ability to sit at a computer for long stretches of time. Researchers have begun to explore using an automated transcription process using digital recordings and voice recognition software (VRS). While VRS has improved in recent years, it is not yet available to the general public in a format that can recognize more than one recorded voice. This article outlines a strategy used to circumvent this problem and improve the speed and ease of transcription. The equipment and the Voice Transcription Technique used are outlined, as well as suggestions for future technological advances in transcription.

### Keywords

Transcription, Voice-Recognition Software, Qualitative Data, and Data Preparation

### Creative Commons License



This work is licensed under a [Creative Commons Attribution-NonCommercial-Share Alike 4.0 License](https://creativecommons.org/licenses/by-nc-sa/4.0/).

# **The Voice Transcription Technique: Use of Voice Recognition Software to Transcribe Digital Interview Data in Qualitative Research**

**Jennifer L. Matheson**

Colorado State University, Fort Collins, Colorado

---

*Transcribing interview data is a time-consuming task that most qualitative researchers dislike. Transcribing is even more difficult for people with physical limitations because traditional transcribing requires manual dexterity and the ability to sit at a computer for long stretches of time. Researchers have begun to explore using an automated transcription process using digital recordings and voice recognition software (VRS). While VRS has improved in recent years, it is not yet available to the general public in a format that can recognize more than one recorded voice. This article outlines a strategy used to circumvent this problem and improve the speed and ease of transcription. The equipment and the Voice Transcription Technique used are outlined, as well as suggestions for future technological advances in transcription. Key Words: Transcription, Voice-Recognition Software, Qualitative Data, and Data Preparation*

---

## **The Context for the Development of the Voice Transcription Technique**

Having had conversations at professional conferences with other qualitative researchers who have managed transcription tasks, I discovered that many of us were attempting to simplify this task using voice recognition software. Some qualitative researchers, like me, had actually attempted to use the software with our recorded interviews, only to discover the result of the transcription was a useless jumble of nonsensical words. All of the researchers I have talked to cited the same barrier, that currently-available voice recognition software does not recognize more than one voice, therefore appearing to be useless in automating the transcription of digital interview data where there are at least two voices recorded. Having learned that other researchers were also trying to use the existing technology to simplify the transcription process, I searched the published literature only to discover there was no one publishing about these techniques. At the same time, I was in the process of conducting my dissertation research. Having conducted qualitative research for many years, I was intrigued by how rapidly qualitative research was advancing in terms of qualitative analysis software, yet how slowly transcription technology was improving. I felt sure that there had to be a way to use the existing technology to simplify the process of transcribing. At the same time, I was developing a significant case of carpal tunnel syndrome from years of typing on poorly designed computer keyboards. After investigating all of the voice recognition software packages available and spending hours imagining a way to use what was available to accomplish my task of transcribing, I developed the following technique to

use during my dissertation research. This process can be used by researchers to lessen the time and physical effort of traditional transcription. In addition to making some of the more mundane qualitative research tasks easier, I hope to improve the use of technology for transcription in qualitative research using this Voice Transcription Technique. By extending the use of technology in this way, this article will encourage others to use creative ways to incorporate technological advancements to continue to improve qualitative methods.

### **The Problem with Transcription**

Qualitative researchers often generate huge quantities of text from interviews, focus groups, observations, or document examinations. Transcription is one step qualitative researchers across the world take on their way to managing and analyzing recorded data. Transcription is also a crucial aspect of the data management process for anyone conducting advanced data analysis or using computer aided qualitative data analysis software (CAQDAS) such as Atlas.ti or NVivo. As crucial as it is, however, many researchers grapple with the task of transcribing their recorded data, experiencing it as a tiresome, lengthy, and challenging process that takes specialized skills, patience, and physical ability (Agar, 1996; Lapadat & Lindsay, 1999; Tilley, 2003). One article examining students' transcription experiences quoted some of their comments as, "the transcription process is intensive and tough" and "the whole process of doing the transcription is lonely and tiring" (Roulston, deMarrais, & Lewis, 2003, p. 657). It is a task that is so often lamented by researchers that experts such as Patton (2002) have gone so far as to publish tips for ways researchers conducting qualitative interviews can help "to keep transcribers sane" (p. 382). Many researchers pass the task of transcription to a clerical assistant, research assistant, graduate student, or a professional transcriptionist because it is a difficult, time-consuming task. For some, it is an issue of lack of time and interest in transcription, while for others it is due to physical limitations. At least one prominent qualitative scholar Ron Chenail (2005) at Nova Southeastern University in Florida, has called for the development and publication of methods to automate transcription.

In recent decades, technological advancements have made many aspects of data collection, management, and analysis easier and faster. Researchers have increasingly relied on improving technology to simplify their most challenging research tasks such as recording field notes and managing large quantities of codes for analysis. While technology has aided qualitative researchers in many ways, innovative technology is unavailable to simplify the transcription of recorded data with multiple voices (i.e., focus groups and interviews). For those of us who are used to utilizing new technology to improve our research, we have found frustratingly few options for simplifying the time-consuming, physically-taxing job of transcription.

More simplified transcription techniques would potentially lead to more researchers doing their own transcriptions. Some researchers believe that transcribing one's own qualitative interview data allows the researcher to grow closer and more familiar with the data (Lapidat & Lindsey, 1999; Tilley, 2003; Wengraf, 2001). It is one of many ways to build in additional theoretical sensitivity during the research process (Strauss & Corbin, 1990). Sometimes referred to as the "researcher-transcriber," the

researcher who chooses to transcribe her/his own data “takes the opportunity to listen carefully and think deeply about the recorded voices and the interview context, using sensory and other memory” (Park & Zeanah, 2005, p. 246). It provides a unique opportunity for interviewers to critique their own work and potentially improve upon their interviewing technique (Anderson & Jack, 1991). Writing memos and journaling are also important aspects of the qualitative research process, and this tends to be more concentrated and fruitful during transcription (Wengraf). Researchers may find it easier to write memos of their thoughts, feelings, reactions, and analytic assumptions during transcription than when the actual data collection occurs, thereby giving them the opportunity to see the parts of the data as pieces of the greater whole. A richer set of memos potentially leads to better insights and a broader set of theoretical questions to explore during analysis. Listening to a recording of an interview provides a flood of thoughts and memories that are not ever-available, and should be recorded as theoretical memos before those memories fade (Wengraf). These thoughts and impressions can be as important in later phases of the analysis and write-up as were the verbatim transcriptions.

The “researcher-transcriber” role allows the interviewer multiple opportunities to hear the interviewee’s words, pauses, silences, and non-verbal expressions such as sighs or crying. Researchers can listen carefully during the transcribing process, so as not to speak or focus on what questions to ask next, as they must during data collection. It is a unique opportunity to be focused on the data without being distracted by the process of data collection. It is also an opportunity to pick up on any ways in which the interviewer can improve or change questions for future data collection. For all of these reasons, many qualitative researchers believe that transcribing one’s own data is highly desirable (Park & Zeanah, 2005; Wengraf, 2001). While it may be desirable, it is a task that some researchers find to be a chore (Agar, 1996; Rettie, 2005) and an extremely time-consuming process (Gibson, Callery, Campbell, Hall, & Richards, 2005). The purpose of this article, therefore, is to outline a new strategy used to maximize the benefits of transcription, while minimizing the negative aspects, leading to a quicker, more efficient, and less tedious outcome. In addition, this article aims to help people who have issues such as carpal tunnel syndrome to be able to transcribe interview data themselves.

### **Voice Recognition Software**

Voice recognition software (VRS) is computer software that automatically transcribes digital voice recordings without the need for typing. It has been available to the general public since the early 1980s, with the most recent versions touting up to a 98% accuracy rate (Al-Aynati & Chorneyko, 2003), a rate higher than many human transcriptionists can boast. In addition, the software has improved in the past 2 decades from one that understands one word at a time with pauses in between to one that understands continuous speech. The newer versions of the software also have extensive vocabularies in multiple languages and dialects that can be altered as needed. In addition, the software packages have a much-improved capacity to be trained by the user to learn new words to improve the quality, speed, and accuracy of the transcription. If there is a word that the software consistently misunderstands, for example, the user can stop, enter the training mode of the program, and help the software learn the word to correctly

recognize the word in the future. VRS also learns by repetition, so the more the user uses the software, the better it comprehends the words and speech patterns of the user. Overall, VRS continues to improve in terms of its accuracy and faster response times (Beirne, 2001).

While the technology of VRS has improved significantly over the past 2 decades, it is designed to be used by one voice at a time. The program is capable of understanding more than one voice, but it cannot access its knowledge of multiple voices simultaneously. While simultaneous multiple voice recognition technology is available in places like governmental and military intelligence communities, it is cost prohibitive for the average researcher and is not yet available commercially. One additional drawback is that the VRS technology available commercially is improving so quickly that as soon as the software is trained, it may be time to replace it with a package that is advanced in all of its functions. While this may be true for some software, most of the better known VRS companies (i.e., Nuance's Dragon Systems or IBM's ViaVoice) upgrade their software such that it does not require new training.

VRS has been used for decades to aid people with physical, developmental, and learning disabilities in working and communicating more effectively. Published studies show that this technology has been used effectively for this purpose for over 30 years (De La Paz, 1999; Kerchner & Kistinger 1984; Roberts, 1999). Lodato (2005) writes about the benefits and complications of using VRS as a woman with multiple sclerosis. She describes the ability of the software to pick up and attempt to spell even the subtlest sounds such as heavy sighs. She points out that in order to be successful using this technology, it takes patience, trial-and-error, and word training within the software program. Those caveats aside, Lodato states that VRS opens up a world of possibilities for people with disabilities.

Only one published article has been found describing the successful use of VRS for transcribing qualitative data in research. Noticing that many researchers were struggling with transcription, Park and Zeanah (2005) conducted tests of two ways to use VRS in transcribing multiple voices in recorded interviews, in addition to the traditional manual form of transcription, to see which was more efficient and less physically demanding. They found that the preferred method was what they call "listen and repeat." This involved the researcher

training the program to his or her voice, listening to the tape recording of the interview/discussion using a conventional transcribing machine and headphones to stop and start the recording, then repeating segments of what was on the tape into the digital microphone and thence to the computer. (p. 246)

The researcher also trained the VRS using his/her own voice and ran it during the interview, attempting to accurately transcribe his/her voice regardless of its accuracy in recognizing the interviewee's voice. One problem with this technique was that VRS learns as it is used. If it is learning incorrect interpretations of words of a second voice that it was not trained on, it is not learning well and becomes a much less effective research tool.

Speed and accuracy are both important considerations in using VRS for transcription. According to Park and Zeanah (2005), their “listen and repeat” technique took roughly the same time as it would for a competent typist, an average of 12 hours. In addition, the authors found that the VRS worked well with people with different accents. Since VRS is trained to understand each person’s unique pattern of speech and dialect, it is versatile enough to be used by people speaking with varied accents and in different languages. The software even includes English versions for multiple forms of English speech, including “American.” On the other hand, it is most important that the speaker’s speech is clear and consistent to maximize the technique even though it can adjust to different accents (Park & Zeanah).

Besides the one article looking at the use of VRS for transcribing qualitative research data, there are a few others who have published on the use of VRS for basic, one voice transcription (see Anderson, 1998; Lee, 2004; Maloney & Paolisso, 2001; Pearson, 2005). Pearson provides a number of key points for people using VRS for optimal efficiency. Tips include ensuring the type and quality of the computer hardware, buying the best external hardware such as the microphone, and speaking clearly and slowly into the VRS system. The current article extends the work of others who have attempted to bridge recorded interview data and automated voice transcription techniques.

### **The Voice Transcription Technique**

The technique outlined in this article is similar to the one outlined by Park and Zeanah (2005) that they called “listen and repeat.” I suspect I was developing this technique at the same time as Park and Zeanah, but I used my technique during a research project, tested it on 13 actual in-depth interviews, and I provide explicit details on how to use it and what equipment is needed to replicate the technique. By following the steps of the Voice Transcription Technique outlined below, qualitative researchers can speed up their transcripts and relieve the physical stress often experienced by classic transcription.

I used this technique during my dissertation research, which was a qualitative study of women in substance abuse recovery. This article focuses only on the equipment and the techniques used for the transcription process during my research project. Since the equipment is such a crucial aspect of the successful use of this technique, I will begin by outlining the type of equipment needed followed by specific instructions on how to use the technique.

### **Equipment**

Three key pieces of hardware and three types of computer software are needed to produce the desired results outlined in this article. Researchers need a computer, a digital recorder, headphones, a microphone, and batteries for the recorder. The software needed includes VRS, transcription software, and word processing software. I used an HP Pavillion laptop computer with the Microsoft® Windows® XP™ operating system. It is important to have a fast computer processor, and I used a 1 GHz processor with 512 MB RAM and 1 GB of free hard drive space in order for the software to perform adequately. Be sure to examine the required computer specifications for the software you are buying to ensure compatibility. I used Microsoft® Office Word 2003 as my word processing

package. Besides the basic computing hardware and software, there were other items needed to accomplish this technique. While the brand names of the equipment are relatively unimportant, the hardware and software products shown in Table 1 are the updated (2007) versions of what I used in 2005.

Table 1

*List of Hardware and Software*

<b>Type of Equipment</b>	<b>Brand</b>	<b>Model</b>	<b>Approx. Cost as of 2008</b>	<b>Specifications</b>
Digital Voice Recorder and MP-3 Player	Sony	ICD-SX57	\$140	256 MB Flash Memory, up to 90 hours of recording time
Voice Recognition Software	Nuance	Dragon Naturally Speaking Version 9	\$99	Standard Version
Transcription Software	Sony	Digital Voice Editor	Included in cost of Sony Digital Voice Recorder	Version 2
Headphones/Microphone	Plantronics®	.Audio™ 345 Behind-the-Head Stereo Headset	\$30	Full-range stereo sound, noise-cancelling microphone

Some of the hardware and software available on the market today are better than others, though most of them will work for the Voice Transcription Technique. In addition, technology is constantly upgrading and improving, therefore it is worth spending time investigating how the equipment recommended above has improved before investing in new equipment. Always buy equipment that is easy to return in case certain pieces are not compatible with one another or not performing well during transcription. Maloney and Paolisso (2001) provide more information that may be useful in deciding what equipment to buy.

### **Preparation for Using the Transcription Technique**

Users of this technique must have some basic familiarity and comfort with personal computers and associated software. If users do not have a basic level of skill with computers, it is advisable not to attempt this technique and to opt instead for the traditional form of transcription.



Before I describe the actual transcription technique, I will provide a few guidelines to follow during data collection to provide optimal results. First, for those who have never recorded their interviews using a digital recorder, it is important to test the equipment you are using thoroughly before using it for data collection. Study the user's manual and spend time getting to know the many functions of the recorder that may be unfamiliar. There are often many settings that one must decide upon before the recording begins. Most digital recorders have optional settings for the quality of the recording, which coincide with the size of the resulting digital files. Highest quality stereo recordings provide the best results, but they take up the most space, while lowest quality mono recordings take up the least. Purchasing a more expensive recorder with a much larger memory capacity would allow the user to use the highest quality setting and still record many hours of interviews.

After each interview, I copied the digital recording from the recorder to my computer hard drive and saved a back-up copy on a flash drive in case anything happened to the original version. No participant names were used in naming the files, and their names were not used during the recording to protect confidentiality. After I transferred the digital files to my computer, I permanently deleted them from the digital recorder to preserve confidentiality. Even though most digital recorders have a locking function that further ensures confidentiality, it is important to delete recordings from portable devices to preserve confidentiality of the data. Deleting the files from the recorder also frees up space for subsequent recordings.

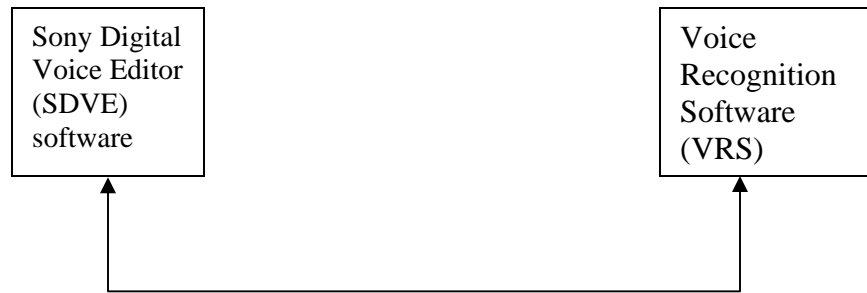
The digital voice recorder I used came with software called Sony Digital Voice Editor (SDVE) that had to be loaded onto my computer. The same is true for the VRS. After saving all of the digital recordings onto my computer, I trained the VRS to recognize my voice. Each VRS has its own directions for training the software to understand the transcriptionist's voice. Follow the directions of the VRS you are using. Most software takes less than an hour to initially train, though VRS "learns" as the user continues using it. Its accuracy is at its lowest point in the first few hours of use and improves its accuracy the longer it is used by the same user. In addition, the user can spend additional time with the VRS, teaching it new words and improving the quality over time. I recommend that users explore the various ways to improve the quality of the VRS by using the manuals or tutorials provided with the software. Due to time constraints, I did not spend more than an hour training the VRS beyond the basic training, since I did not know how well the software would work initially. I found the training to be interesting and fun, providing some insight into how the software works.

After I trained the transcription software for voice recognition, I opened both the VRS and the SDVE software so that both were active and visible. Be sure to "resize" each of the two software packages so that both can be seen and utilized at the same time on your computer screen. Because these pieces of software use a considerable amount of memory, it is important not to have superfluous software running at the same time (specifically if you are using a laptop with less than the previously recommended processor, memory, and hard drive). Computers with more memory may be able to handle other software running simultaneously, but it is not advisable in order to avoid any processing interruptions during transcription. I found that the software ran more smoothly and more quickly when I had no additional software running while the VRS and SDVE software were running.

It is important to mention here the distinct yet collaborative roles of the two software packages; the VRS and the SDVE software. While they work together in seamless harmony, they have distinct roles, as summarized in Table 2. The SDVE software is responsible for managing the digital recording and feeding the sound into the headphones. This software helps to manage the speed, volume, and tone of the digital recording, as well as the playback length and ability to start, stop, pause, fast-forward, rewind, and restart as needed. These functions are exactly like traditional transcription machines, only this software allows the user to do all functions through his/her computer. The SDVE software also made it possible to skip entire sections if needed. Since the data were digital, it made quick work of maneuvering through the recordings using this software. Being able to adjust the amount of rewind when a segment needed to be replayed for clarity, as well as the adjustable tone, helps the user keep up with the dialogue and hear it at its optimal clarity. The VRS, on the other hand, is responsible for interpreting the user’s spoken word, then automatically typing it into a text file without the user having to touch the keyboard. It stops transcribing when the user stops talking (or as soon as it finishes transcribing everything up to the point where the user stopped speaking) and starts up again as soon as it hears the user’s voice again. There is a natural lag in the VRS’ ability to keep up with the natural speed of the user’s voice, especially if it is set at its most sensitive setting. While this may prove frustrating for some users initially, it can also be seen is a benefit in that it allows the user to take a break from time to time while the VRS catches up.

Table 2

*Comparing the Functions of SDVE Software and VRS*



Function of the SDVE Software	Function of the VRS
<ul style="list-style-type: none"> <li>• Playback the digital recording through the headphones</li> <li>• Manages the speed, tone, and playback of the digital recording</li> </ul>	<ul style="list-style-type: none"> <li>• Receive researcher’s verbal recitation of the recording through the microphone</li> <li>• Transcribe the user’s words into a text file</li> </ul>

Before beginning the actual transcription, it is important that the user place the integrated headphone/microphone on her/his head, so that the headset is comfortable and secure on each ear and the microphone is very close to the corner of the mouth (but not touching the mouth). Carefully read the headset instructions to ensure proper positioning.

### **The Transcription Technique**

Once the software is ready and the headset is properly positioned, open a new, blank file into the VRS and name it according to that interviewee's ID number. This is where the actual transcription will be saved, in a new, blank document in the VRS. Next, the corresponding digital recording that had previously been downloaded and saved on the computer should be opened into the SDVE software. Begin playback of the digital recording by using the computer mouse to click the play button on the SDVE screen. The interview should be heard clearly through the headphones, and volume, speed, and tone adjustments can be made using the SDVE software for optimal sound. As the words of the dialogue come through the headset, the user should repeat into the microphone what he/she hears through the headset. The user's spoken words are simultaneously transcribed by the VRS, and text begins to appear on the blank document of the VRS. The user may stop speaking to take a break, to go back and make corrections (either verbally or with the keyboard), or when the recording ends. As the VRS transcribes, the user can see it attempting to find the best matches for words it hears.

In addition to understanding the words spoken into the microphone, the VRS also understands certain words or combinations of words as commands. This is known as command mode. Because recent advances in VRS provide for "command mode" and "dictate mode," it understands that some words the user says are commands asking the software to fulfill a function instead of transcribe. It takes very little time to learn the few commands needed to quickly and efficiently transcribe the interview into text. Such is the case with the words "new paragraph" and "colon" that are helpful to learn. If instead of typing a colon the user wants the software to transcribe the word "colon," the user would say into the microphone, "spell out c-o-l-o-n," saying each separate letter of the word "colon" one by one. Hearing the words "spell out" before a set of letters tells the VRS to transcribe the word colon, not the punctuation mark of a colon. When the phrase "spell out" is used in any case, the VRS always defaults to command mode and spells the letters that the user recites. Once the user stops spelling out single letters and the VRS finds the correct word to transcribe, the VRS returns to "dictate mode" and the user can continue speaking normally. This feature is useful for many tasks including navigating around the document, editing, and helping to train the software.

One important element needed in any transcript is a way to distinguish the interviewer from the interviewee(s) as each begins to speak. The VRS manual provides instructions on how to accomplish this. For example, if the user wants to start a new paragraph for the words of the interviewer, the user would speak the following command: "New paragraph interviewer colon." No pauses are needed in saying these words. This command tells the VRS to move to the next free line in the text file and type the word "interviewer" followed by a colon, and then continue transcribing. Use this command every time the speaker of the text changes from the interviewer to the interviewee and

back again. Also use this same command to tell the VRS that a new section or paragraph is required.

This process continues until the entire recording is transcribed. After each transcript is complete, the user should replay the recording again to make additional edits to ensure the most accurate transcription possible, not unlike what might be expected of a traditional transcriptionist. This not only helps to improve accuracy of the transcript, but for those researchers who are doing their own transcription, it also provides another chance to hear the interview, become closer to the data, and record any final memos or journal entries for analysis. It is perhaps easier for researchers with no major manual disabilities to make corrections to the transcript by hand instead of using the VRS. This process is fast, and the user will rarely need to stop the recording as she/he reviews the transcript a final time for accuracy. There will probably be corrections to be made, but they may be so few and dispersed that the user will rarely have to stop the recording.

A person with serious manual disabilities, on the other hand, will need to use the VRS to make final corrections. One feature that will be useful to researchers using the VRS for corrections is the command “go to.” An example of this would be if the user finds the word “break” instead of the correct spelling “bake.” This command prompts the VRS to search for the next occurrence of the word that follows the command, in this case the word “break.” This “go to” command is like the “find” function in a word processor program. When the VRS finds the word, the user uses the command “delete break” to delete the word. The user then speaks the correct word “bake” to instruct the VRS to retype the correct word. If the VRS does not type the word “bake” correctly a second time, the user can either go into training mode and train the software on this word, or use the “spell out” command followed by the letters b-a-k-e as described earlier. After all mistakes have been corrected, the transcript is complete. Remember to save the document frequently throughout the process to ensure work will not be lost. It is also recommended to save a backup copy of each transcribed text file to ensure the files are not lost permanently through either user or technology failures.

### **Digital Transcribing Tips**

There are a number of important tips to remember as researchers embark upon this Voice Transcription Technique. As mentioned earlier, it is imperative that the speech of the user be very clear and consistent. While there is some normal variance in a person’s voice from one day to the next, the software does not perform as well when this variation occurs. Every effort should be made to complete a transcript in one day if possible. Another tip is to practice speaking into the microphone and allowing the VRS to transcribe in a separate document as a warm-up. After a few minutes, when it appears the user’s voice and the software are in sync, the actual interview transcription can begin.

Another important hint is to ensure that one’s surroundings are very quiet, private, and free of extraneous noise. This is both so that the VRS does not pick up additional noises during the transcription, but also because this technique takes considerable concentration, especially early on in the process. Additionally, it can be annoying for those around the user to hear the one-sided, long, monotone recitation of a digitally recorded interview. Similarly, confidentiality is always an important consideration of

well-executed qualitative research. It is important that people who should not overhear the actual interview not be within earshot of the user during the transcription process.

Besides controlling the transcribing environment, it is also important that the user carefully place the equipment for optimal use. The microphone must be very close to the user's mouth without being so close as to pick up breathing noises that will interfere with the automated transcription. The headset-microphone allows for a lot of flexibility since the microphone is not mounted to a desk, and it allows the user's head to move freely since it is attached to his or her head. The computer and keyboard also need to be at a level and distance from the user that is not unusually tiring for one's head, neck, shoulders, and back. Finally, be sure the screen is close enough and the type is large enough to be able to be clearly seen.

### **Discussion**

By combining transcription software, VRS, and an integrated headphone/microphone headset, I developed an innovative Voice Transcription Technique that utilizes and tests recent advances in technology. It allows qualitative researchers and people with manual disabilities to use their voices to transcribe multiple-voice interview data. There are a number of important advantages to using this innovative technique for transcribing digitally recorded data. Digital recordings are more reliable and the quality is advanced compared to traditional tape recordings. Digital recordings are more easily transported, transferred, and are at less risk of being destroyed after they have been saved. There is some risk of these recordings being inadvertently deleted before they have been transferred to a computer, but there is a locking function on most digital recorders that make the user go through multiple steps before a recording can be erased. Recording over an existing interview is equally difficult in that it takes multiple steps to record over a previous recording. In addition, digital recordings are much improved in terms of sound clarity and quality. This enhances the transcription because the user is better able to determine exactly what is said in a multiple voice recording. Other nuanced speech and noises such as sighs, mumbling, laughter, and inflection are also much easier to hear in digital voice recordings.

Another major advantage to using this technology is the ease for people with difficulty typing, such as those with manual disabilities, severe arthritis, or carpal tunnel syndrome. Transcribing can be physically very difficult with wrist, back, and eye strain always a factor. This is a similar finding to that of Park and Zeanah (2005) who found, "a disability preventing easy or prolonged keyboard use, lack of money to pay others and, in some cases, a slow typing speed, provided the motivation" (p. 248) to use this technique successfully. Transcribing many interviews can also be psychologically draining after hours of such a monotonous activity. Quality of transcripts suffers when both the physical and mental strains become too much. This method of transcription is less physically and mentally taxing, and makes transcription less of a chore. It is also easy to transcribe in a variety of settings since all of the equipment is portable. For even easier transportability, I would recommend a cordless integrated headphone/microphone headset. It allows for easier movement around the computer and prevents additional cords from interfering with accessibility to all hardware.

While this process of transcribing proved to be faster and less physically and mentally taxing than traditional transcription, the overall transcription speed improved significantly over time. This is partly due to the VRS's ability to learn and improve its accuracy over time. It is also because the user is better able to pace his or her speech, articulate more carefully in a way that helps the VRS respond, and learn the VRS commands. This is a similar finding to that of Park and Zeanah (2005) who found after only a few tries "that the time taken using VRS was roughly equal to the time taken by a competent typist (i.e., 4 to 5 hours for an hour-long tape of moderately good quality)" (p. 248).

Most of the results of using the technology outlined in this article are similar to the advantages found by Park and Zeanah (2005). These include the ability to listen carefully to the interviews, adding memos during transcription, the ability to transcribe multiple voices, the ability of those with disabilities to use it, and the increased speed of transcription compared to traditional methods. Cost is another benefit, both because the equipment needed for this technique is no more than that which a traditional transcription project would cost, not to mention not having to hire a research assistant or professional transcriptionist to perform the work. Additional advantages of this technique are: the ease, transportability, and security of using digital recordings; the advantage of mobility using an integrated headset; and the lack of physical and mental exhaustion often experienced with traditional transcription. A few disadvantages include the need for computer competence, the need for time for training, and only a modest savings of time in transcription for beginning users of the technique. Once the user becomes more proficient with the technique, the time savings becomes evident. Park and Zeanah pointed out a few additional disadvantages including initial frustrations with the VRS and one's own performance, voice fatigue, and difficulty detecting small remaining errors when high levels of accuracy are achieved.

While most qualitative researchers are prepared for the eventual arrival of affordable and available VRS that is capable of understanding more than one voice, those who follow the guidelines laid out by this article should find the technique a significant improvement to traditional transcription. Those who continue to develop innovative ways to make lengthy and physically taxing research tasks easier should publish them as quickly as possible to benefit all of those struggling with similar issues. Finally, even when affordable and technologically advanced VRS arrives, there will be a list of other research challenges to deal with that could benefit from the publication of creative approaches.

## References

- Agar, M. H. (1996). *The professional stranger: An informal introduction to ethnography*. San Diego, CA: Academic Press.
- Al-Aynati, M. M., & Chorneyko, K. A. (2003). Comparison of voice-automated transcription and human transcription in generating pathology reports. *Archives of Pathology & Laboratory Medicine*, 127(6), 721-725.
- Anderson, J. (1998). Transcribing with voice recognition software: A new tool for qualitative researchers. *Qualitative Health Research*, 8(5), 718-723.

- Anderson, K., & Jack, D. (1991). Learning to listen: Interview techniques and analysis. In S. Gluck & D. Patai (Eds.), *Women's words* (pp. 11-26). New York: Routledge.
- Beirne, M. (2001). Finding voice recognition software that works. *Review of Ophthalmology*, 8(2), 27-34.
- Chenail, R. J. (2005). Future directions for qualitative methods. In D. H. Sprenkle & F. P. Piercy, *Research methods in family therapy* (2nd ed., pp. 191-210). New York: Guilford.
- De La Paz, S. (1999). Composing via dictation and speech recognition systems: Compensatory technology for students with learning disabilities. *Learning Disability Quarterly*, 22(3), 173-182.
- Gibson, W., Callery, P., Campbell, M., Hall, A., & Richards, D. (2005). The digital revolution in qualitative research: Working with digital audio data through Atlas.ti. *Sociological Research Online*, 10(1), 1-10.
- Kerchner, L. B., & Kistinger, B. J. (1984). Language processing/word processing: Written expression, computers, and learning disabled students. *Learning Disability Quarterly*, 7(4), 329-335.
- Lapidat, J., & Lindsey, A. (1999). Transcription in research and practice: From standardization of technique to interpretive positionings. *Qualitative Inquiry*, 5, 64-86.
- Lee, R. M. (2004). Recording technologies and the interview in sociology, 1920-2000. *Sociology*, 38(5), 869-889.
- Lodato, J. (2005). Advances in voice recognition. *The Futurist*, 39(1), 7-8.
- Maloney, R. S., & Paolisso, M. (2001). What can digital audio data do for you? *Field Methods*, 13(1), 88-96.
- Park, J., & Zeanah, A. E. (2005). An evaluation of voice recognition software for use in interview-based research: A research note. *Qualitative Research*, 5(2), 245-251.
- Patton, M. Q. (2002). *Qualitative research & evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.
- Pearson, J. (2005, October 7). Fire the transcriptionist--but keep dictating. *Medical Economics*. Retrieved August 3, 2006, from <http://www.memag.com/memag/article/articleDetail.jsp?id=182797&searchString=speech%20recognition>
- Rettie, R. (2005, Winter). Exploiting freely available software for social research. *Social Research Update*, 48. Retrieved on September 4, 2006, from <http://sru.soc.surrey.ac.uk/SRU48.html>
- Roberts, R. J. (1999). Use of computer dictation by students with learning disabilities. *Dissertation Abstracts International*, 60, 9-A. (UMI No. 9946502)
- Roulston, K., deMarrais, K., & Lewis, J. B. (2003). Learning to interview in the social sciences. *Qualitative Inquiry*, 9(4), 643-668.
- Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: Sage.
- Tilley, S. A. (2003). "Challenging" research practices: Turning a critical lens on the work of transcription. *Qualitative Inquiry*, 9(5), 750-773.
- Wengraf, T. (2001) *Qualitative research interviewing: Semi-structured, biographical and narrative methods*. London: Sage.

---

**Author Note**

Jennifer L. Matheson, Ph.D., is an Assistant Professor in the Department of Human Development and Family Studies at Colorado State University. She has degrees in sociology and Marriage & Family Therapy. Dr. Matheson's research focuses on addictions and treatment. Her address is Colorado State University, 1570 Campus Delivery, Fort Collins, CO, 80523; Telephone: (970) 491-7472; Email: jemathes@cahs.colostate.edu

Copyright 2007: Jennifer L. Matheson and Nova Southeastern University

**Article Citation**

Matheson, J. L. (2007). The voice transcription technique: Use of voice recognition software to transcribe digital interview data in qualitative research. *The Qualitative Report*, 12(4), 547-560. Retrieved from <http://www.nova.edu/ssss/QR/QR12-4/matheson.pdf>

---